REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
21-08-2014	Final Report	1-Sep-2013 - 31-May-2014

4. TITLE AND SUBTITLE
Final Report: Identification of Gene Networks: An Approach Based on Mathematical Modeling

5a. CONTRACT NUMBER
W911NF-13-1-0384

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Desmond S. Lun

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Rutgers, The State University of New Jersey
Rutgers, The State University of New Jersey Camden C
311 N. 5th Street
Camden, NJ 08102 -1515

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
63317-MA-II.2

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The reverse engineering of transcriptional regulatory networks is one of the grand challenges of systems biology. In this project, we sought to develop a mathematical theory to determine a minimal set of experimental measurements needed to reverse engineer a transcriptional regulatory network. We developed a theoretically near-optimal reverse engineering method called the sensitivity method. We showed through computational experiments that, compared to predominant existing approaches, the sensitivity method leads to vastly reduced experimental cost and greater accuracy. On a 100-gene network, the experimental cost is reduced by an order of magnitude, with the level of

15. SUBJECT TERMS
Gene regulation, gene networks, reverse engineering

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Desmond Lun
UU	UU	UU	UU		19b. TELEPHONE NUMBER 856-225-6094

Report Title

Final Report: Identification of Gene Networks: An Approach Based on Mathematical Modeling

ABSTRACT

The reverse engineering of transcriptional regulatory networks is one of the grand challenges of systems biology. In this project, we sought to develop a mathematical theory to determine a minimal set of experimental measurements needed to reverse engineer a transcriptional regulatory network. We developed a theoretically near-optimal reverse engineering method called the sensitivity method. We showed through computational experiments that, compared to predominant existing approaches, the sensitivity method leads to vastly reduced experimental cost and greater accuracy. On a 100-gene network, the experimental cost is reduced by an order of magnitude, with the level of reduction increasing as the size of the network increases. We applied the sensitivity method to a five-gene subnetwork of Escherichia coli and obtained promising preliminary experimental results.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

S. Sharma, D. Lun, J.-C. Birget, D. Hong, and A. Wirth. A Theoretical Approach to Gene Network Identification. Presented at 2013 RECOMB/ISCB Conference on Regulatory and Systems Genomics, with DREAM Challenges, November 8-12, 2013, Toronto, Canada.

Number of Presentations: 1.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received	Paper	
03/28/2014	1.00	Jean-Camille Birget, Desmond S. Lun, Sweta Sharma, Anthony Wirth, Dawei Hong. A Theoretical Approach to Gene Network Identification, IEEE TRANSACTIONS ON INFORMATION THEORY (03 2014)

TOTAL: **1**

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Sweta Sharma	0.00	
Pooi Kuhn Ip	0.56	
FTE Equivalent:	**0.56**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Evgeni Nikolaev	0.43
FTE Equivalent:	**0.43**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Desmond S. Lun	0.06	
Jean-Camille Birget	0.00	
Dawei Hong	0.00	
Anthony Wirth	0.00	
FTE Equivalent:	**0.06**	
Total Number:	**4**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Anatoliy Lane	0.12	Computer Science
FTE Equivalent:	**0.12**	
Total Number:	**1**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See Attachment

Technology Transfer

Final Report: Identification of Gene Networks: An Approach Based on Mathematical Modeling

Statement of the problem studied

The gene regulatory network of an organism is a key component in understanding its biology and the way in which it responds to external stimuli. Thus, significant attention has been focused in the biological community on determining gene regulatory networks using high-throughput -omics data—so-called "reverse engineering" of gene networks. Despite extensive efforts, however, there are no feasible methods for accurately determining large-scale gene networks. We seek, through a combination of theoretical and experimental approaches, to develop an accurate, feasible reverse engineering method that can be applied to large-scale gene networks.

Summary of the most important results

1. We developed a theoretically near-optimal reverse engineering method that we call the sensitivity method. Using an ordinary differential equation (ODE) model of gene networks developed for the DREAM (Dialogue on Reverse Engineering Assessment and Methods) competition[4], we computationally simulated gene networks and their behavior under experimental perturbations. We showed that, on networks satisfying this ODE model, the sensitivity method reverse engineers gene networks at vastly reduced experimental cost and with greater accuracy. On a 100-gene network, the experimental cost is reduced by an order of magnitude, with the level of reduction increasing as the size of the network increases (see Appendix 1)[2].
2. We applied the sensitivity method to a five-gene subnetwork of *Escherichia coli*: *ompR*, *flhC*, *flhD*, *flgA*, and *flgC*. The regulatory interactions among these genes have been previously discovered and are part of the regulatory interaction database RegulonDB[7]. We aimed to show that the sensitivity method recovers the known regulatory interactions. Our preliminary results show that the sensitivity method can be used to partially recover the subnetwork, and with some modifications to reduce biological replicate variability and some additional measurements, we expect to be able to fully recover the subnetwork. In addition, our method discovered a novel interaction among *flhC* and *flhD* that is consistent with small RNA regulation (see Appendix 2).

Bibliography

1 Balaji, S., Babu, M. M., Iyer, L. M., Luscombe, N. M. & Aravind, L. Comprehensive Analysis of Combinatorial Regulation using the Transcriptional Regulatory Network of Yeast. *J. Mol. Biol.* **360**, 213-227, doi:10.1016/j.jmb.2006.04.029 (2006).

2 Birget, J.-C., Lun, D. S., Sharma, S., Wirth, A. & Hong, D. A Theoretical Approach to Gene Network Identification. (Submitted).

3 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640-6645 (2000).

4 de Jong, H. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* **9**, 67-103, doi:10.1089/10665270252833208 (2002).
5 Marbach, D., Prill, R. J., Schaffter, T., Mattiussi, C., Floreano, D. & Stolovitzky, G. Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. USA* **107**, 6286-6291, doi:10.1073/pnas.0913357107 (2010).
6 Prill, R. J., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., Clarke, N. D., Altan-Bonnet, G. & Stolovitzky, G. Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges. *PLoS ONE* **5**, e9202, doi:10.1371/journal.pone.0009202 (2010).
7 Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J. S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernández, S., Alquicira-Hernández, K., López-Fuentes, A., Porrón-Sotelo, L., Huerta, A. M., Bonavides-Martínez, C., Balderas-Martínez, Y. I., Pannier, L., Olvera, M., Labastida, A., Jiménez-Jacinto, V., Vega-Alvarado, L., del Moral-Chávez, V., Hernández-Alvarez, A., Morett, E. & Collado-Vides, J. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* **41**, D203-D213, doi:10.1093/nar/gks1201 (2013).
8 Schaffter, T., Marbach, D. & Floreano, D. GeneNetWeaver: *in silico* benchmark generation and performance profiling of network inference methods. *Bioinformatics* **27**, 2263-2270, doi:10.1093/bioinformatics/btr373 (2011).
9 Shin, S. & Park, C. Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. *J. Bacteriol.* **177**, 4696-4702 (1995).
10 Stafford, G. P., Ogi, T. & Hughes, C. Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator $FlhD_2C_2$. *Microbiology* **151**, 1779-1788, doi:10.1099/mic.0.27879-0 (2005).

Appendices

Appendix 1: Results from computational experiments

The sensitivity method was computationally tested on *in silico* networks with varying assumptions and sizes. Networks were generated using GeneNetWeaver 3.1[8], the software used to generate networks for DREAM3, 4, and 5 challenges. These were randomly generated from global interaction networks of all experimentally validated regulatory interactions in either *Escherichia coli*[7] or *Saccharomyces cerevisiae*[1].

A dynamical model was constructed using the ODE system described by Marbach et al.[5] and Schaffter et al.[8], and parameters were randomly generated. Measurement noise was simulated by adding Gaussian noise with varying standard deviation to the steady-state expression level of each gene. Expression was normalized so values laid in the interval [0,1]. For noisy networks, changes in expression were detected using a two-sample t-test. Network identification accuracy for noisy networks was assessed by calculating the average area under the precision-recall curve (AUPR). Precision is the ratio of the number of correctly identified interactions over the total number of identified interactions. Recall is the ratio of the number of correctly identified interactions over the total number of interactions. Precision-recall curves were traced by scanning the significance level, s, used for the t-test, starting from s sufficiently large that the recall is 1, with geometrically decaying values until recall reached 0.

We assessed the performance of the sensitivity method for identification of noiseless

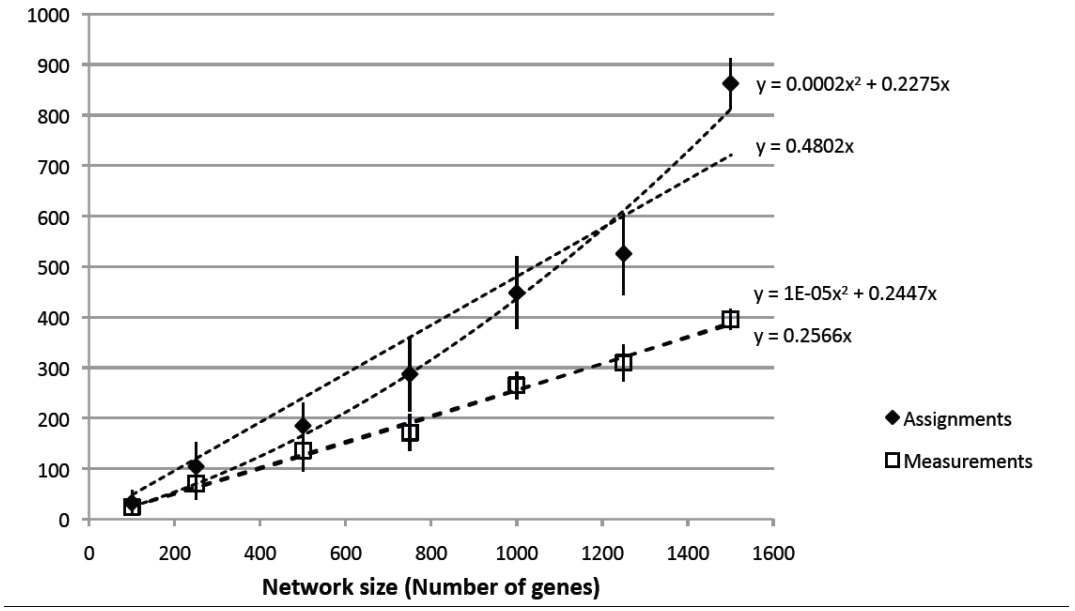

Figure 1. Average assignment and measurement complexity for acyclic *E. coli* networks without noise. Averages were calculated over 10 randomly generated networks. Error bars indicate standard deviation.

acyclic networks. To obtain acyclic networks, networks were generated using GeneNetWeaver and strongly connected components (SCCs) were detected and broken using depth-first search and removing back-edges. Network sizes ranged from 100 to 1,500 for *E. coli* and 500 to 4,000 for *S. cerevisiae*. For each species, the largest network generated is smaller than the number of known genes because of the lack of regulatory information for many genes. The number of assignments and measurements required for *E. coli* networks appears to increase linearly with increasing network size (Figure 1). For *S. cerevisiae* networks, assignments and measurements seem to increase faster than linearly with increasing network size (Figure 2). This faster than linear increase in complexity may be explained by the tendency for *S. cerevisiae* networks to have longer diameters and nodes with, on average, higher degrees compared to *E. coli* networks.

Data in Figure 1 and Figure 2 were fit using linear and quadratic regression. Fits were used to extrapolate results and estimate complexity costs for identification of hypothetical full-sized acyclic *E. coli* and *S. cerevisiae* networks of 4,501 and 6,607 genes respectively. Using linear fitting, we estimate that approximately 2,100 assignments and 1,100 measurements would be needed for identification of the full *E. coli* network, compared to approximately 5,000 assignments and 1,300 measurements using quadratic fitting. For identification of the full *S. cerevisiae* network, linear estimates require approximately 5,600 assignments and 2,900 measurements compared to quadratic estimates of 12,600 assignments and 8,300 measurements. Because we do not know how complexity costs change as networks approach their full size, we take the linear and quadratic estimates for complexity as approximate lower and upper bounds respectively.

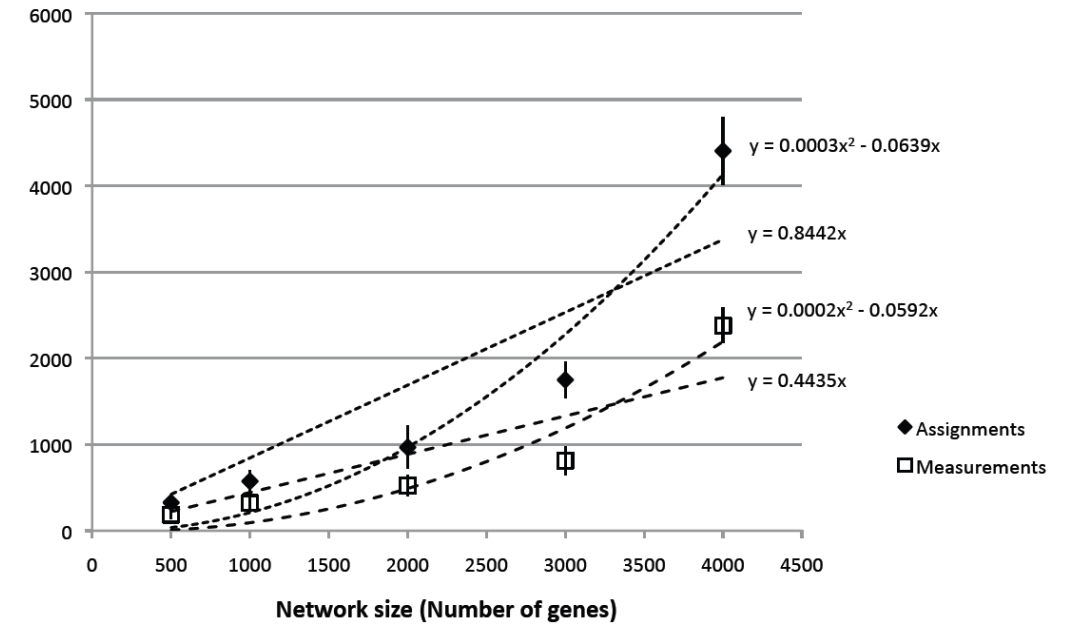

Figure 2. Average assignment and measurement complexity for acyclic *S. cerevisiae* networks without noise. Averages were calculated over 10 randomly generated networks. Error bars indicate standard deviation.

Assignments and measurements were determined for noiseless cyclic *E. coli* networks with ten replicates of each size, from 250 to 1,500 (Figure 3). We found that less than 5% of the genes in each network were part of a SCC (Figure 4). The number of assignments and measurements required for identification of these networks, excluding their SCCs, seem to increase linearly with increasing network size. Thus, using linear extrapolation, we estimate that a hypothetical full-size *E. coli* network of 4,501 genes containing cycles would require, on average, approximately 2,100 assignments and

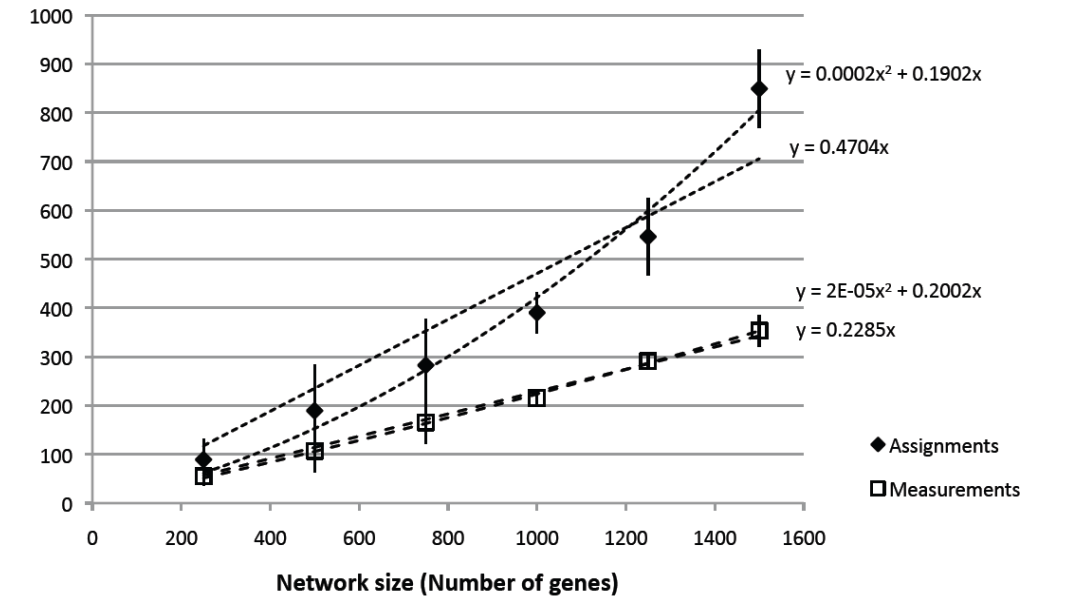

Figure 3. Average assignment and measurement complexity for cyclic *E. coli* networks without noise. Averages were calculated over 10 randomly generated networks. Error bars indicate standard deviation.

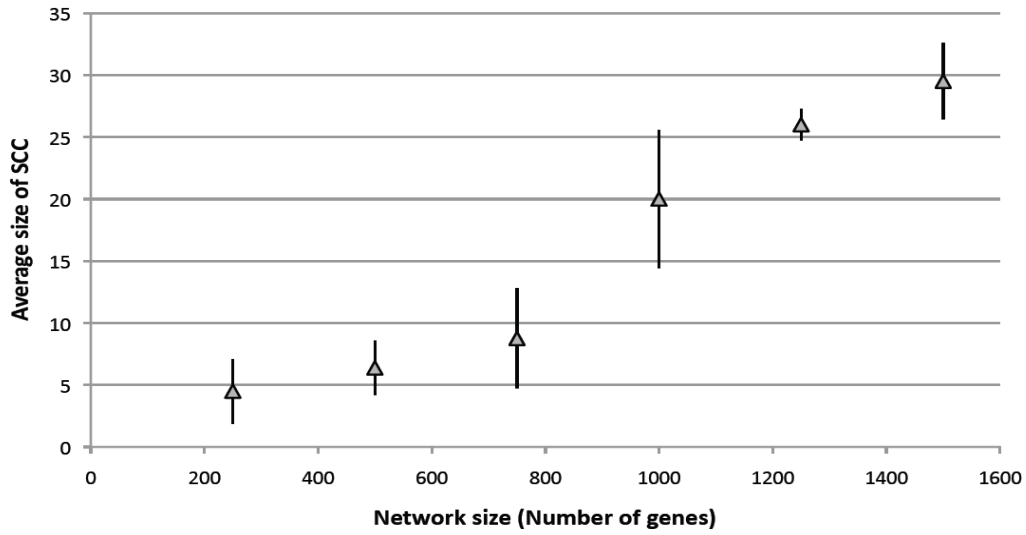

Figure 4. Average size of SCCs for cyclic E. coli networks without noise. Averages were calculated over ten randomly generated networks. Error bars indicate standard deviation.

1,000 measurements to be identified up to the SCCs. With quadratic extrapolation, on average, 4,900 assignments and 1,300 measurements would be required.

Algorithm performance for noisy *E. coli* networks of 100 and 500 genes was assessed with respect to varying number of replicate measurements and varying levels of additive Gaussian noise ranging from standard deviation of 0.01 to 0.1 (Figure 5). For the 100-gene network, a minimum of five replicate measurements is needed to achieve an average AUPR of at least 0.95 for noise standard deviation of 0.05. For the 500-gene network, however, an average AUPR of only 0.76 is achieved even after 64 replicates with noise standard deviation of 0.01 (Figure 6). The minimum and maximum AUPR

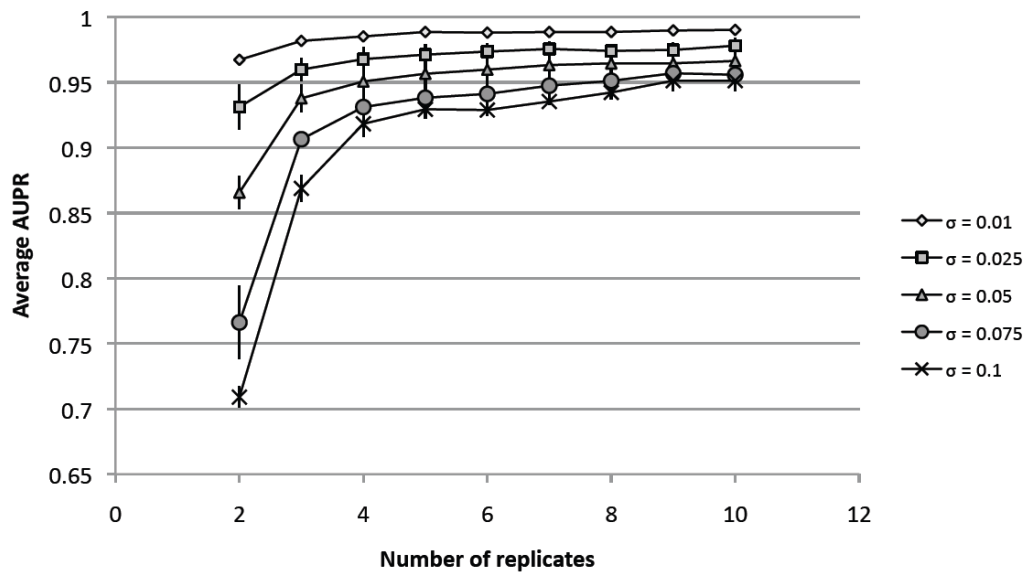

Figure 5. Average AUPR for varying number of measurement replicates and noise standard deviation (σ) in 100-gene *E. coli* network simulations. Averages were calculated over five randomly generated networks. Error bars indicate standard deviation.

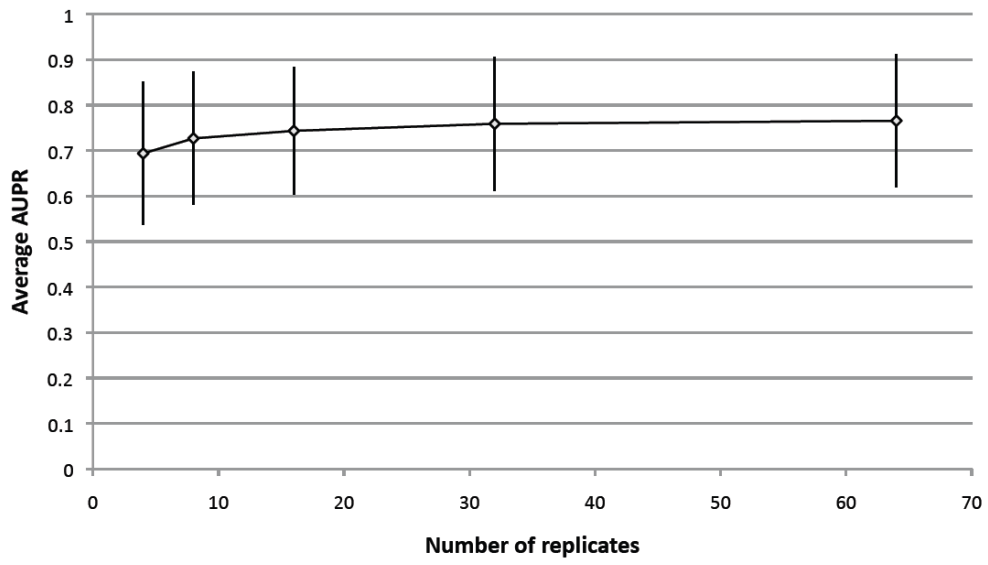

Figure 6. Average AUPR for varying number of measurement replicates in 500-gene *E. coli* network simulations with noise standard deviation 0.01. Averages were calculated over 29 randomly generated networks. Error bars indicate standard deviation.

among the 29 networks tested is 0.39 and 0.98, respectively. Such high variability may be related to error propagation. It seems that, for larger networks, there is a significant probability that GeneNetWeaver will generate a gene with multiple regulators where the effect of one specific regulator by itself is small (e.g., a close approximation to a binary AND gate or OR gate is achieved). Errors associated with the detection of this regulatory interaction then appear to propagate in our algorithm. It is unclear whether such regulation functions occur in real, natural systems or whether they are an artifact of the parameter generation distributions used by GeneNetWeaver.

We compare our results to those from the DREAM3 competition[5,6]. For the network inference challenge, teams were supplied with three types of data: steady-state expression levels of wild-type and single gene knockout strains, steady state expression levels of single gene knockdowns, and 46 time courses of expression level trajectories for wild-type strains with different multi-factorial perturbations. The trajectories were sampled at 21 time-points. Gaussian noise with 0.05 standard deviation was added to all data points to simulate measurement error. For the 100-gene network challenge, these data represent 246 assignments and 1,167 measurements if we consider each multi-factorial perturbation only as a single assignment. Even so, the best-performing team's algorithm achieved an average AUPR of 0.75. In comparison, we were able to achieve 0.95 AUPR on 100-gene networks with approximately 32 assignments and 120 measurements on average.

Appendix 2: Experimental results on five-gene *E. coli* network

To experimentally validate the model, we chose an acyclic sub-network in *E. coli* MG1655 consisting of five non-essential genes: outer-membrane protein regulator ompR, master regulators of flagellar synthesis genes *flhD* and *flhC*, and flagellar biosynthesis genes f*lgA* and *flgC*. The arc set for this network is {(*ompR*, *flhD*), (*ompR*, *flhC*), (f*lhD*, *flgA*), (*flhD*, *flgC*), (f*lhC*, *flgA*), (*flhC*, *flgC*)} (Figure 7). There is strong

experimental evidence for each of these interactions[9,10]. In theory, the sensitivity network for this set would contain transitivity arcs (*ompR*, *flgA*) and (*ompR*, *flgC*). The cutset for both transitivity arcs is {*flhD*, *flhC*}. Single gene knockouts for the transcription factors *ompR*, *flhD*, and *flhC* were created from MG1655 using the λ Red mediated recombination system developed by Datsenko et al.[3] (primers listed in Table 1). To turn off the cutset genes, a double knockout of *flhD* and *flhC* was constructed and transformed with plasmid pZS2-123 cloned with *flhDC* downstream of the *lacZ* promoter . The *flhDC* fragment was amplified from wild-type MG1655. pZS2-123 was restriction digested using BamH1-HF restriction enzyme and cloning was performed via Gibson Assembly. The strains are summarized in Table 2.

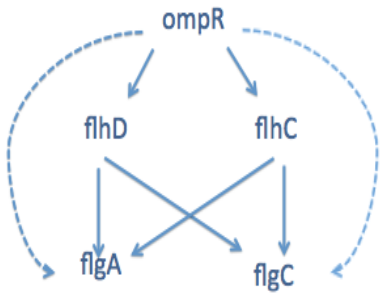

Figure 7. *E. coli* regulatory sub-network containing genes *ompR*, *flhD*, *flhC*, *flgA*, and *flgC*. Solid edges represent interactions for which there is strong experimental evidence. Dashed edges represent potential sensitivity edges (i.e. indirect interactions where the target is affected by a perturbation at the source, but there is no direct, physical interaction between the two).

To experimentally obtain the sensitivity network, cultures from S1, S2, and S3 strains as well as wild-type were grown in triplicate to $OD600_{nm}$ ~1 in M9 media supplemented with 1mM $MgSO_4$ and .2mM $CaCl_2$. RNA was collected and DNA digested using the

Table 1. Sequences of primers used to create knockout strains.

Primer	Sequence 5' to 3'
ompR P1	GAATACACGCTTACAAATTGTTGCGAACCTTTGGGAGTACAAACAATGCAGT GTAGGCTGGAGCTGCTTC
ompR P2	TGAACTTCGTGGCGAGAAGCGCAATCGCCTCATGCTTTAGAGCCGTCCGGATT CCGGGGATCCGTCGACC
flhC P1	TGTTAATCAGCCTGAAGAAGCGCTGCGCAAGAAAAGGGCCTGATCATGAGGT GTAGGCTGGAGCTGCTTC
flhC P2	TGCTGGAATGTTGCGCCTCACCGTATCAGTTAAACAGCCTGTACTCTCTGATT CCGGGGATCCGTCGACC
flhD P1	GGTGAAACCGCATAAAAATAAAGTTGGTTATTCTGGGTGGGAATAATGCAGT GTAGGCTGGAGCTGCTTC
flhD P2	TTCCTGAACAATGCTTTTTTCACTCATGATCAGGCCCTTTTCTTGCGCAGATT CCGGGGATCCGTCGACC
flgA P1	TGCGGACAGGTACAATTCACGTTGTAGAAATGGCTGGGGGCGAAAATGCTGT GTAGGCTGGAGCTGCTTC
flgA P2	TCGGCAAGGGACGGGTAATCTTTAACAGCTTACAGGTTTATAAGAATATTAT TCCGGGGATCCGTCGACC
flgC P1	CAAATCAAAGGCATGATGAACGTTTTACAGAGCGGAAATTAACGGATGGCGT GTAGGCTGGAGCTGCTTC
flgC P2	GGTTACCGCAATGGACATAGCTTTCTCCTTTATTGACCGAGCGTAAGGGTATT CCGGGGATCCGTCGACC

Qiagen RNeasy Mini Kit and RNase-free DNase set. qPCR was performed for two experimental replicates of each RNA sample. Approximately 50ng total RNA and 200nM primers were used with the qScript™ One-Step SYBR Green qRT-PCR Kit. Primers were designed using the PrimerQuest tool from the Integrated DNA Technologies website with amplicon length ranging from 100-110 bp. For thermocycling, the standard qPCR cycling protocol specified in the qScript manual was used over 40 cycles.

Table 2. Summary of strains for validation study.

Strain	Genotype
S1	MG1655Δ*ompR*
S2	MG1655Δ*flhD*
S3	MG1655Δ*flhC*
S4	MG1655Δ*flhDC* pZS2-123
S5	MG1655Δ*flhDC*Δ*ompR* pZS2-123

Technical duplicates for which the standard deviation of the Ct value exceeded 2 were excluded from analysis. The remaining Ct values were used to perform two-sample t-tests to determine differential expression of each gene between strains. p-values for the t-tests are summarized in Table 3. We see that the largest p-values are obtained for *ompR* in Δ*flhD* and Δ*flhC*, which is what we expect, since *flhD* and *flhC* do not regulate *ompR*. Smaller p-values are obtained for all other interactions, which is what we expect from the network in Figure 7, except for *flhC* in Δ*flhD* and *flhD* in Δ*flhC*. This suggests that *flhC* regulates *flhD* and vice versa. Although there is no evidence of direct regulatory

Table 3. Summary of qPCR results. p-values are for two-sample t-test comparing Ct values of each gene between wild-type (WT) and knockout strains.

Strain	gene	mean Ct	std. dev	p-value
WT	*ompR*	20.55	0.84	
	flhD	35.92	0.34	
	flhC	28.70	0.04	
	flgA	28.03	0.98	
	flgC	27.91	0.28	
Δ*ompR*	*ompR*	31.75	1.18	0.0002
	flhD	34.60	0.25	0.0480
	flhC	27.73	0.93	0.2548
	flgA	27.45	0.64	0.4731
	flgC	26.64	1.13	0.2364
Δ*flhD*	*ompR*	20.16	0.95	0.6237
	flhD	40.00	0.00	0.0002
	flhC	25.69	0.79	0.0330
	flgA	26.13	0.57	0.1419
	flgC	26.91	1.29	0.3969
Δ*flhC*	*ompR*	20.88	0.34	0.5643
	flhD	33.58	0.53	0.0345
	flhC	38.75	1.77	0.0153
	flgA	26.70	0.14	0.0871
	flgC	27.69	0.34	0.5134

interaction between these two genes, the two genes are co-regulated by the same small RNA, which would explain this observation. Although the data in Table 3 show the correct trend, the p-value separation among non-interacting and interacting gene pairs is rather small. In follow-up experiments, we plan to adjust our qPCR assay to additionally measure expression of reference genes to which the expression of all other genes can be normalized. This will allow distinction of differences in expression that are due to differential regulation and those due to noise. Candidate reference genes for *E. coli* MG1655 include glucan biosynthesis protein G (*mdoG*) and 16S ribosomal RNA (*rrsB*). In this manner, we expect to reduce the variability among biological replicates and, in so doing, increase the p-value discrimination between interacting and non-interacting genes.

Finally, we see in our data that knockout of *ompR* affects the expression of *flgA* and *flgC* even though there is no direct regulation of *flgA* or *flgC* by *ompR*. This effect is consistent with the indirect interactions that we expect. In the sensitivity method, we discriminate indirect interactions from direct interactions by measuring the expression of *flgA* and *flgC* under a Δ*flhC* Δ*flhD flhC+ flhD+* mutant (i.e. a mutant where *flhC* and *flhD* have been knocked out and complemented) and a Δ*ompR* Δ*flhC* Δ*flhD flhC+ flhD+* mutant. Because there is no direct regulation of *flgA* and *flgC* by *ompR*, we expect to see no significant difference in their expression between these two mutants.